DOCUMENT RESUME

ED 076 597 TM 002 576

AUTHOR Browne, Michael W.
TITLE Generalized Least Squares Estimators in the Analysis of Covariance Structures.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RB-73-1
PUB DATE Jan 73
NOTE 38p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Analysis of Covariance; *Factor Analysis; Matrices; Measurement Techniques; *Models; Statistical Analysis; Statistical Data; Systems Analysis; Technical Reports

ABSTRACT

This paper concerns situations in which a p x p covariance matrix is a function of an unknown q x 1 parameter vector y-sub-o. Notation is defined in the second section, and some algebraic results used in subsequent sections are given. Section 3 deals with asymptotic properties of generalized least squares (G.L.S.) estimators of y-sub-o. Section 4 concerns methods for obtaining estimates of parameters in certain linear covariance structures. (Author/KM)

FORM R510 PRINTED IN U.S.A.

ED 076597

TM 002 576

RB-73-1

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

RESEARCH BULLETIN

GENERALIZED LEAST SQUARES ESTIMATORS IN THE ANALYSIS OF COVARIANCE STRUCTURES

Michael W. Browne

National Institute for Personnel Research (South Africa)

and

Educational Testing Service

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the author. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
January 1973

Generalized Least Squares Estimators in the Analysis of Covariance Structures

Summary

Let S represent the usual unbiased estimator of a covariance matrix, Σ_o, whose elements are functions of a parameter vector $\underset{\sim}{\gamma}_o$: $\Sigma_o = \Sigma(\underset{\sim}{\gamma}_o)$. A ~~generalized least squares (G.L.S.)~~ estimate, $\hat{\underset{\sim}{\gamma}}$, of $\underset{\sim}{\gamma}_o$ may be obtained by minimizing $tr[\{S - \Sigma(\underset{\sim}{\gamma})\}V]^2$ where V is some positive definite matrix. Asymptotic properties of the G.L.S. estimators are investigated assuming only that $\Sigma(\underset{\sim}{\gamma})$ satisfies certain regularity conditions and that the limiting distribution of S is multivariate normal with specified parameters. The estimator of $\underset{\sim}{\gamma}_o$ which is obtained by maximizing the Wishart likelihood function (M.W.L. estimator) is shown to be a member of the class of G.L.S. estimators with minimum asymptotic variances. When $\Sigma(\underset{\sim}{\gamma})$ is linear in $\underset{\sim}{\gamma}$, a G.L.S. estimator which converges stochastically to the M.W.L. estimator involves far less computation. Methods for calculating estimates of $\underset{\sim}{\gamma}_o$, estimates of the dispersion matrix of $\hat{\underset{\sim}{\gamma}}$, and test statistics, are given for certain linear models.

Some key words: Covariance structures; Generalized least squares; Asymptotic distributions.

Generalized Least Squares Estimators in the

Analysis of Covariance Structures

1. Introduction

This paper will be concerned with situations where a $p \times p$ covariance matrix, Σ_o, is a function of an unknown $q \times 1$ parameter vector γ_o:

$$\Sigma_o = \Sigma(\gamma_o) \quad . \tag{1}$$

Suppose that the p component vectors x_k, $k = 1,2 \ldots n + 1$, are independently and identically distributed with mean μ_o and covariance matrix Σ_o. Let S represent the usual unbiased estimator of Σ_o obtained from the x_k. It has been common practice to assume a multivariate normal distribution for x_k or a Wishart distribution for S, and employ maximum likelihood estimators of γ_o. Nonlinear structures (e.g., Jöreskog, 1970a) and linear structures (e.g., Bock & Bargmann, 1966; Anderson, 1969, 1970) have been investigated. Provided that μ_o is unstructured, maximum likelihood estimators of γ_o based on a multivariate normal distribution for $x_1 \ldots x_{n+1}$, or on a Wishart distribution for S, are functions of S only and differ only by a scaling factor, $n/(n + 1)$. The choice of maximum likelihood estimators is possibly due to their asymptotic efficiency and associated likelihood ratio test. Considering a particular nonlinear covariance structure, the unrestricted factor analysis model, Jöreskog & Goldberger (1972) have shown that a certain generalized least squares estimator also is asymptotically efficient and that a corresponding weighted residual sum of squares statistic converges stochastically to the likelihood ratio statistic.

This paper considers estimators of γ_o which are functions of S where

$$\mathcal{E}\{s_{ij}\} = \sigma_{oij} \quad . \tag{2}$$

The only assumptions about the distribution of elements of S concern the asymptotic distribution as $n \to \infty$ which is to be the multivariate normal distribution with means given by (2) and covariances

$$\mathrm{Cov}(s_{ij}, s_{gh}) = n^{-1}(\sigma_{oig}\sigma_{ojh} + \sigma_{oih}\sigma_{ojg}) \quad . \tag{3}$$

This requires only that all fourth order cumulants of the distribution of the x_k are zero (cf. Cramér, 1946, pp. 365-366; Kendall & Stuart, 1969, p. 321). The results to be given apply to, but are not confined to, the situation where the x_k have a multivariate normal distribution and S has a Wishart distribution.

Section 3 will be concerned with asymptotic properties of generalized least squares (G.L.S.) estimators of γ_o. No specific form will be assumed for the covariance structure model. Results will apply to all models which satisfy certain regularity conditions. Although S may not necessarily have a Wishart distribution one may still obtain estimates by maximizing the Wishart likelihood function. These "M.W.L." estimators will be shown to have the asymptotic properties of the class G.L.S. estimators with minimum asymptotic variances.

When the covariance structure is linear, G.L.S. estimates may be expressed in closed form and are more easily calculated than the M.W.L.

estimates. Section 4 will be concerned with methods for obtaining estimates of parameters in certain linear covariance structures.

The next section defines notation and gives some algebraic results which will be used in subsequent sections.

2. Notation and Preliminary Algebraic Results

The column vector formed from elements of a $p \times p$ matrix, S, taken columnwise will be denoted by $\mathrm{Vec}(S)$ or by the corresponding small letter underlined.

i.e., $\mathrm{Vec}'(S) = \underline{s}' = s_{11}, s_{21}, s_{31}, \ldots, s_{12}, s_{22}, s_{32} \cdots s_{13}, s_{23}, s_{33} \cdots s_{pp}$.

Double subscripts, ij, are used to denote elements of this vector, the first subscript always being nested within the second. Double subscripts will also be used to represent rows or columns of certain matrices. For example, a typical element of the direct product $A \otimes B$ will be denoted by $[A \otimes B]_{ij,gh}$ where

$$[A \otimes B]_{ij,gh} = a_{jh} b_{ig} \quad . \tag{4}$$

Using this expression it is easily shown that

$$(A \otimes B)\underline{s} = \mathrm{Vec}(B\,S\,A') \tag{5}$$

if A and B are of order $m \times p$ and S is of order $p \times p$.

The column vector formed from the elements above and including the diagonal of a symmetric matrix, S, taken columnwise, will be denoted by $\underset{\sim}{s}$.

i.e. $\underset{\sim}{s}' = s_{11}; s_{12}, s_{22}; s_{13}, s_{23}, s_{33}; \ldots\ s_{pp}$.

Again, double subscripts, ij , are used to denote elements of this vector, the first being nested within the second and not exceeding the second.

As the $p \times p$ matrix S is symmetric, the $p(p+1)/2 \times 1$ vector $\underset{\sim}{s}$ may be expressed in terms of the $p^2 \times 1$ vector $\underline{s}$:

$$\underset{\sim}{s} = K_p' \underline{s} \tag{6}$$

where K_p is of order $p^2 \times p\{p+1\}/2$ with typical element

$$[K_p]_{ij,gh} = 2^{-1}(\delta_{ig}\delta_{gh} + \delta_{ih}\delta_{jg}) \quad , \quad i \leq p \ , \quad j \leq p \ ; \quad g \leq h \leq p$$

and δ_{ij} represents Kronecker's delta. Therefore,

$$[K_p]_{ii,ii} = 1$$

$$[K_p]_{ij,ij} = [K_p]_{ij,ji} = 1/2 \qquad i \neq j$$

$$[K_p]_{ij,gh} = 0 \quad \text{if} \quad ij \neq gh \quad \text{and} \quad ij \neq hg \quad .$$

A left inverse of K_p is

$$K_p^- = (K_p'K_p)^{-1}K_p' \tag{7}$$

which is of order $p\{p+1\}/2 \times p^2$ with typical element

$$[K_p^-]_{gh,ij} = (2 - \delta_{gh})[K_p]_{ij,gh} \quad , \quad i \leq p \ , \ j \leq p \ ; \ g \leq h \leq p$$

$$= 1 \quad \text{if} \quad ij = gh \quad \text{or} \quad ij = hg$$

$$= 0 \quad \text{otherwise.}$$

This matrix may be used to express $\underline{s}$ in terms of $\underset{\sim}{s}$:

$$\underline{s} = K_p^{-\prime}\underset{\sim}{s} \quad . \tag{8}$$

Let M_p represent the $p^2 \times p^2$ symmetric idempotent matrix

$$\begin{aligned} M_p &= K_p(K_p'K_p)^{-1}K_p' \\ &= K_pK_p^- \end{aligned} \tag{9}$$

with typical element

$$[M_p]_{ij,gh} = 2^{-1}(\delta_{ig}\delta_{jh} + \delta_{ih}\delta_{jg}) \quad , \quad i \le p \ , \ j \le p \ , \ g \le p \ , \ h \le p \ .$$

This matrix has an interesting property. If A is of order $p \times m$, then

$$M_p(A \otimes A) = (A \otimes A)M_m \quad . \tag{10}$$

Other properties are:

$$M_pK_p = K_p \quad , \tag{11}$$

and

$$M_p\underline{s} = \underline{s} \quad . \tag{12}$$

The inverse of the matrix $K_p'(W \otimes W)K_p$, where W is nonsingular of order $p \times p$, is

$$\{K_p'(W \otimes W)K_p\}^{-1} = K_p^-(W^{-1} \otimes W^{-1})K_p^{-\prime} \quad . \tag{13}$$

This result may be verified by multiplication using (9), (10), and (11) and the inversion rule of direct products (e.g., Searle, 1966, p. 216):

$$K_p'(W \otimes W)\, M_p(W^{-1} \otimes W^{-1})K_p^{-\prime} = K_p'M_p(W \otimes W)(W \otimes W)^{-1}K_p^{-\prime}$$

$$= I \quad .$$

Let the column vector formed from the diagonal elements of the matrix S be denoted by either $\mathrm{diag}(S)$ or $\underline{\underline{s}}$. The $p^2 \times p$ matrix H_p, with typical element

$$[H_p]_{ij,g} = \delta_{ig}\delta_{jg} \quad , \quad i \leq p \; , \; j \leq p \; , \; g \leq p$$

$$= 1 \quad \text{if} \quad i = j = g$$

$$= 0 \quad \text{otherwise}$$

may be used to select $\underline{\underline{s}}$ from $\underline{s}$:

$$\mathrm{diag}(S) = \underline{\underline{s}} = H_p'\underline{s} \quad . \tag{14}$$

Let $V*W$ represent the term by term product of V and W with typical element $[V*W]_{ij} = [V]_{ij}[W]_{ij}$. Since

$$V*W = H_p'(V \otimes W)H_p \quad , \tag{15}$$

$V*W$ is positive semidefinite if V and W are positive semidefinite.

In subsequent sections it will frequently be convenient to express a quadratic or bilinear form involving a direct product as a trace using:

$$\underline{x}'(V \otimes W)\underline{y} = \mathrm{tr}[XVY'W'] \tag{16}$$

where $\underline{x} = \mathrm{Vec}(X)$ and $\underline{y} = \mathrm{Vec}(Y)$.

We shall regard the $q \times 1$ vector $\underline{\gamma}$ as a mathematical variable which can assume values $\underline{\gamma}_o$ and $\hat{\underline{\gamma}}$, where $\hat{\underline{\gamma}}$ is an estimate of $\underline{\gamma}_o$. $\Sigma = \Sigma(\underline{\gamma})$

will be regarded as a matrix function of $\underline{\gamma}$. When matrix derivatives are given, the equality of the functions $\sigma_{ij}(\underline{\gamma})$ and $\sigma_{ji}(\underline{\gamma})$ will always be taken into account. Matrices of partial derivatives such as $\frac{\partial\Sigma(\underline{\gamma})}{\partial\gamma_i}$ and $\frac{\partial^2\Sigma(\underline{\gamma})}{\partial\gamma_i\partial\gamma_j}$ will therefore be symmetric. $\hat{\Sigma}$, $\frac{\partial\hat{\Sigma}}{\partial\gamma_i}$ and $\frac{\partial^2\hat{\Sigma}}{\partial\gamma_i\partial\gamma_j}$ will stand for $\Sigma(\hat{\underline{\gamma}})$, $\left.\frac{\partial\Sigma(\underline{\gamma})}{\partial\gamma_i}\right|_{\underline{\gamma}=\hat{\underline{\gamma}}}$ and $\left.\frac{\partial^2\Sigma(\underline{\gamma})}{\partial\gamma_i\partial\gamma_j}\right|_{\underline{\gamma}=\hat{\underline{\gamma}}}$ respectively. A similar convention will be employed when $\underline{\gamma} = \underline{\gamma}_o$.

3. Generalized Least Squares Estimators

The model given in (1) may be expressed in the equivalent form

$$\mathcal{E}(\underline{s}) = \underline{\sigma}_o = \underline{\sigma}(\underline{\gamma}_o) \quad . \tag{17}$$

We shall assume throughout that this model satisfies the following regularity conditions:

(a) All $\sigma_{ij}(\underline{\gamma})$ and all partial derivatives of the first three orders with respect to elements of $\underline{\gamma}$ are continuous and bounded in a neighborhood of $\underline{\gamma} = \underline{\gamma}_o$.

(b) The $p^2 \times q$ matrix

$$\Delta = \left.\frac{\partial\underline{\sigma}(\underline{\gamma})}{\partial\underline{\gamma}'}\right|_{\underline{\gamma}=\underline{\gamma}_o} \tag{18}$$

is of full column rank.

(c) $\underline{\gamma}_o$ is identified, i.e., $\Sigma(\underline{\gamma}_1) = \Sigma(\underline{\gamma}_o)$ implies that $\underline{\gamma}_1 = \underline{\gamma}_o$.

(d) $\Sigma(\underline{\gamma}_o)$ is positive definite.

Let us consider the residual quadratic form,

$$\{\underset{\sim}{s} - \underset{\sim}{\sigma}(\underline{\gamma})\}'\{\mathrm{Cov}(\underset{\sim}{s},\underset{\sim}{s}')\}^{-1}\{\underset{\sim}{s} - \underset{\sim}{\sigma}(\underline{\gamma})\} \quad . \tag{19}$$

It follows from the Gauss-Markov theorem that, if $\underset{\sim}{\sigma}(\underline{\gamma})$ is linear in $\underline{\gamma}$, minimization of this residual quadratic form yields the minimum variance unbiased estimator of $\underline{\gamma}_o$. If $\underset{\sim}{\sigma}(\underline{\gamma})$ is nonlinear, the estimator will be asymptotically efficient.

In order to obtain $\{\mathrm{Cov}(\underset{\sim}{s},\underset{\sim}{s}')\}^{-1}$, the matrix of this quadratic form, we use (4) to express (3) as

$$\mathrm{Cov}(s_{ij}, s_{gh}) = n^{-1}\Big(\tfrac{1}{2}\{[\Sigma_o \otimes \Sigma_o]_{ij,gh} + [\Sigma_o \otimes \Sigma_o]_{ji,hg}\}$$

$$+ \tfrac{1}{2}\{[\Sigma_o \otimes \Sigma_o]_{ji,gh} + [\Sigma_o \otimes \Sigma_o]_{ij,hg}\}\Big)$$

so that

$$\mathrm{Cov}(\underset{\sim}{s},\underset{\sim}{s}') = 2n^{-1}K_p'(\Sigma_o \otimes \Sigma_o)K_p \quad . \tag{20}$$

Then, (12) shows that the required inverse is

$$\{\mathrm{Cov}(\underset{\sim}{s},\underset{\sim}{s}')\}^{-1} = 2^{-1}nK_p^{-}(\Sigma_o \otimes \Sigma_o)K_p^{-\prime} \tag{21}$$

so that, with use of (8), the quadratic form (19), which we now denote by $nf(\underline{\gamma}|\Sigma_o^{-1})$, becomes

$$nf(\underline{\gamma}|\Sigma_o^{-1}) = 2^{-1}n\{\underset{\sim}{s} - \underset{\sim}{\sigma}(\underline{\gamma})\}'K_p^{-}(\Sigma_o^{-1} \otimes \Sigma_o^{-1})K_p^{-\prime}\{\underset{\sim}{s} - \underset{\sim}{\sigma}(\underline{\gamma})\}$$

$$= 2^{-1}n\{\underline{s} - \underline{\sigma}(\underline{\gamma})\}'(\Sigma_o^{-1} \otimes \Sigma_o^{-1})\{\underline{s} - \underline{\sigma}(\underline{\gamma})\} \quad . \tag{22}$$

The matrix of this quadratic form is a function of the unknown dispersion matrix Σ_o . We shall therefore replace Σ_o^{-1} by another matrix, V , and consider G.L.S. estimators which result from minimizing

$$f(\gamma|V) = 2^{-1}\{\underset{\sim}{s} - \underset{\sim}{\sigma}(\gamma)\}'(V \otimes V)\{\underset{\sim}{s} - \underset{\sim}{\sigma}(\gamma)\} \tag{23}$$

with respect to γ . The weight matrix, V , will be either a stochastic matrix which converges in probability to a positive definite matrix $\bar{V}$ as $n \to \infty$ or a positive definite constant matrix $(V = \bar{V})$. Consequently the matrix of the quadratic form in (23) is positive definite or converges in probability to a positive definite matrix, $\bar{V} \otimes \bar{V}$. Using (16) this quadratic form may also be expressed as:

$$f(\gamma|V) = 2^{-1}\, \mathrm{tr}[\{S - \Sigma(\gamma)\}V]^2 \quad . \tag{24}$$

We shall examine asymptotic properties of the estimators.

Proposition 1. The G.L.S. estimators are consistent.

Proof. Since γ_o is identified and $\bar{V}$ is positive definite, $\mathrm{tr}[\{\Sigma_o - \Sigma(\gamma)\}\bar{V}]^2$ has its absolute minimum of zero at $\gamma = \gamma_o$. S and V converge stochastically to Σ_o and $\bar{V}$ and $\Sigma(\gamma)$ is bounded in a neighborhood of $\gamma = \gamma_o$. Consequently $\mathrm{tr}[\{S - \Sigma(\gamma)\}V]^2$ converges in probability to $\mathrm{tr}[\{\Sigma_o - \Sigma(\gamma)\}\bar{V}]^2$ uniformly in a neighborhood of $\gamma = \gamma_o$. Since $\mathrm{tr}[\{S - \Sigma(\gamma)\}V]^2$ is continuous in γ , the point $\hat{\gamma}$ where it has its absolute minimum converges stochastically to γ_o . This proof is an adaptation of a proof of Anderson & Rubin (1956, pp. 145-146).||

Proposition 2. The limiting distribution of a G.L.S. estimator, $\hat{\underset{\sim}{\gamma}}$, is multivariate normal with mean vector

$$\mathcal{E}(\hat{\underset{\sim}{\gamma}}) \simeq \underset{\sim}{\gamma}_o \tag{25}$$

and covariance matrix

$$\mathrm{Cov}(\hat{\underset{\sim}{\gamma}}, \hat{\underset{\sim}{\gamma}}') \simeq 2n^{-1}\{\Theta(\bar{V})\}^{-1}\Theta(\bar{V}\Sigma_o\bar{V})\{\Theta(\bar{V})\}^{-1} \tag{26}$$

where $\Theta(\bar{V})$ is a $q \times q$ matrix function of $\bar{V}$ defined by

$$\Theta(\bar{V}) = \Delta'(\bar{V} \otimes \bar{V})\Delta \tag{27}$$

with typical element [c.f. (16), (18)]

$$[\Theta(\bar{V})]_{ij} = \mathrm{tr}\left(\frac{\partial\Sigma_o}{\partial\gamma_i} \bar{V} \frac{\partial\Sigma_o}{\partial\gamma_j} \bar{V}\right) \quad .$$

Proof. Let

$$h(\underset{\sim}{\gamma}|V) = - \frac{\partial f(\underset{\sim}{\gamma}|V)}{\partial\underset{\sim}{\gamma}} = \frac{\partial\underset{\sim}{\sigma}'}{\partial\underset{\sim}{\gamma}} \{V \otimes V\}\{\underset{\sim}{s} - \underset{\sim}{\sigma}(\underset{\sim}{\gamma})\}$$

Using (16), a typical element of this vector may be expressed as

$$h_i(\underset{\sim}{\gamma}|V) = \mathrm{tr}[V\{S - \Sigma(\underset{\sim}{\gamma})\}V \frac{\partial\Sigma}{\partial\gamma_i}] \quad .$$

By Taylor's theorem

$$\underset{\sim}{h}(\hat{\underset{\sim}{\gamma}}|V) = h(\underset{\sim}{\gamma}_o|V) - W(\hat{\underset{\sim}{\gamma}} - \underset{\sim}{\gamma}_o) \tag{28}$$

where

$$[W]_{ij} = -\left.\frac{\partial h_i}{\partial \gamma_j}\right|_{\gamma=\gamma_o} - \frac{1}{2}\sum_{k=1}^{q}(\hat{\gamma}_k - \gamma_{ok})\left.\frac{\partial^2 h_i}{\partial \gamma_j \partial \gamma_k}\right|_{\gamma=\gamma^*}$$

and γ^* lies between γ_o and $\hat{\gamma}$.

Now,

$$\frac{\partial h_i}{\partial \gamma_j} = \mathrm{tr}[V\{S - \Sigma(\gamma)\}V \frac{\partial^2 \Sigma}{\partial \gamma_i \partial \gamma_j}] - \mathrm{tr}[V \frac{\partial \Sigma}{\partial \gamma_i} V \frac{\partial \Sigma}{\partial \gamma_j}] \qquad (29)$$

and

$$\frac{\partial^2 h_i}{\partial \gamma_j \partial \gamma_k} = \mathrm{tr}[V\{S - \Sigma(\gamma)\}V \frac{\partial^3 \Sigma}{\partial \gamma_i \partial \gamma_j \partial \gamma_k} - V \frac{\partial \Sigma}{\partial \gamma_i} V \frac{\partial^2 \Sigma}{\partial \gamma_j \partial \gamma_k}$$

$$- V \frac{\partial \Sigma}{\partial \gamma_j} V \frac{\partial^2 \Sigma}{\partial \gamma_k \partial \gamma_i} - V \frac{\partial \Sigma}{\partial \gamma_k} V \frac{\partial^2 \Sigma}{\partial \gamma_i \partial \gamma_j}] \quad . \qquad (30)$$

Since the elements of $\{S - \Sigma(\gamma_o)\}$ and $(\hat{\gamma} - \gamma_o)$ converge to zero in probability, since the trace functions in (29) and (30) are continuous, and since the partial derivatives are asymptotically bounded in probability it follows that $[W]_{ij}$ converges stochastically to $\mathrm{tr}(\bar{V} \frac{\partial \Sigma_o}{\partial \gamma_i} \bar{V} \frac{\partial \Sigma_o}{\partial \gamma_j})$, or

$$\operatorname*{plim}_{n\to\infty} W = \Delta'(\bar{V} \otimes \bar{V})\Delta = \Theta(\bar{V})$$

as can be seen from (16), (18). This matrix is nonsingular.

Since $h(\hat{\gamma}|V) = 0$, it follows from (28) that

$$\hat{\gamma} = \gamma_o + W^{-1}h(\gamma_o|V)$$

and $\hat{\underset{\sim}{\gamma}}$ is asymptotically equivalent to

$$\ddot{\underset{\sim}{\gamma}} = \{\Theta(\bar{V})\}^{-1}\underline{h}(\underset{\sim}{\gamma}_o|\bar{V})$$

$$= \{\Xi(\bar{V})\}'(\underline{s} - \underline{\sigma}_o) \quad ,$$

where

$$\Xi(\bar{V}) = (\bar{V} \otimes \bar{V})\Delta\{\Theta(\bar{V})\}^{-1} \quad , \tag{31}$$

because

$$\sqrt{n}\,(\hat{\underset{\sim}{\gamma}} - \ddot{\underset{\sim}{\gamma}}) = \{W^{-1} - [\Delta'(\bar{V} \otimes \bar{V})\Delta]^{-1}\}\Delta'(\bar{V} \otimes \bar{V})\{\sqrt{n}\,(\underline{s} - \underline{\sigma}_o)\}$$

$$+ W^{-1}\Delta'\{(V \otimes V) - (\bar{V} \otimes \bar{V})\}\{\sqrt{n}\,(\underline{s} - \underline{\sigma}_o)\} \tag{32}$$

converges in probability to the null vector as $n \to \infty$.

Since $\ddot{\underset{\sim}{\gamma}}$ is a linear function of $\underline{s}$, the limiting distribution of $\ddot{\underset{\sim}{\gamma}}$ and of $\hat{\underset{\sim}{\gamma}}$ is multivariate normal with mean vector

$$\{\Xi(\bar{V})\}'\Delta\underset{\sim}{\gamma}_o = \underset{\sim}{\gamma}_o$$

and dispersion matrix

$$\{\Xi(\bar{V})\}'K_p^{-'}\,\mathrm{Cov}(\underset{\sim}{s},\underset{\sim}{s}')K_p^{-}\{\Xi(\bar{V})\} = 2n^{-1}\{\Xi(\bar{V})\}'M_p\{\Sigma_o \otimes \Sigma_o\}M_p\{\Xi(\bar{V})\} \quad .$$

This dispersion matrix may be expressed in the form of (26) after use of (31), (27), (10), (12), and the fact that each column of Δ is formed from a symmetric matrix, $\partial\Sigma_o/\partial\gamma_j$. ||

All G.L.S. estimators of γ_0, then, are consistent and asymptotically normally distributed. The "best" G.L.S. (B.G.L.S) estimators, in the sense of having minimum asymptotic variances, are obtained by taking V to be some consistent estimator of $\kappa\Sigma_0^{-1}$ where κ is any positive constant.

<u>Proposition 3</u>. The asymptotic dispersion matrix of a G.L.S. estimator, $\hat{\underline{\gamma}}$, is bounded below by $2n^{-1}\{\Theta(\Sigma_0^{-1})\}^{-1}$ in the Loewner sense of inequality (e.g., Beckenbach & Bellman, 1965, p. 86). This bound is attained, and $\hat{\underline{\gamma}}$ is a B.G.L.S. estimator, if $\bar{V} = \kappa\Sigma_0^{-1}$. ($\kappa > 0$)

<u>Proof</u>. $\{\Theta(\bar{V})\}^{-1}\Theta(\bar{V}\Sigma_0\bar{V})\{\Theta(\bar{V})\}^{-1} - \{\Theta(\Sigma_0^{-1})\}^{-1}$

$$= \{\Xi(\bar{V}) - \Xi(\Sigma_0^{-1})\}'\{\Sigma_0 \otimes \Sigma_0\}\{\Xi(\bar{V}) - \Xi(\Sigma_0^{-1})\}$$

$$\geq 0$$

since $\Sigma_0 \otimes \Sigma_0 > 0$. ||

In order to prove asymptotic efficiency of B.G.L.S. estimators we would have to show that the difference between $2n^{-1}\{\Theta(\Sigma_0^{-1})\}^{-1}$ and the inverse information matrix (based on the exact distribution of S) is of the order $o(n^{-1})$. If S has a Wishart distribution, this difference is the null matrix so that all B.G.L.S. estimators are efficient. If we assume only that the limiting distribution of S is multivariate normal with parameters given by (17) and (20), we can say that B.G.L.S. estimators are "efficient in terms of the limiting distribution of S " in the following sense:

Proposition 4. Let Ω denote the information matrix based on the limiting distribution of S . Then

$$\lim_{n\to\infty} n[2n^{-1}\{\Theta(\Sigma_o^{-1})\}^{-1} - \Omega^{-1}] = 0 \quad . \tag{33}$$

Proof. The log of the likelihood function for the limiting multivariate normal distribution of $\underset{\sim}{s}$ is

$$\log L_N = \text{constant} - \frac{1}{2}\{\log|K_p'\{\Sigma(\underline{\gamma}) \otimes \Sigma(\underline{\gamma})\}K_p| + \frac{n}{2}\,\mathrm{tr}[S\{\Sigma(\underline{\gamma})\}^{-1} - I]^2\}$$

with first derivatives,

$$\frac{\partial \log L_N}{\partial\gamma_i} = \frac{n}{2}\,\mathrm{tr}[\Sigma^{-1}(S - \Sigma)\Sigma^{-1}S\Sigma^{-1}\frac{\partial\Sigma}{\partial\gamma_i}] - \frac{(p+1)}{2}\,\mathrm{tr}[\Sigma^{-1}\frac{\partial\Sigma}{\partial\gamma_i}]$$

and second derivatives

$$\frac{\partial^2 \log L_N}{\partial\gamma_i\partial\gamma_j} = -\frac{n}{2}\{\mathrm{tr}[\Sigma^{-1}S\Sigma^{-1}\frac{\partial\Sigma}{\partial\gamma_i}\Sigma^{-1}S\Sigma^{-1}\frac{\partial\Sigma}{\partial\gamma_j} + 2\Sigma^{-1}\frac{\partial\Sigma}{\partial\gamma_i}\Sigma^{-1}(S - \Sigma)\Sigma^{-1}S\Sigma^{-1}\frac{\partial\Sigma}{\partial\gamma_j}$$

$$- \Sigma^{-1}(S - \Sigma)\Sigma^{-1}S\Sigma^{-1}\frac{\partial^2\Sigma}{\partial\gamma_i\partial\gamma_j}] + (p+1)n^{-1}\,\mathrm{tr}[\Sigma^{-1}\frac{\partial^2\Sigma}{\partial\gamma_i\partial\gamma_j}$$

$$- \Sigma^{-1}\frac{\partial\Sigma}{\partial\gamma_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\gamma_j}]\} \quad . \tag{34}$$

Using (16), (17), and (20) it can easily be shown that, if Q_1 and Q_2 are $p \times p$ matrices and $\delta = 0$ or 1 ,

$$\mathcal{E}\,\mathrm{tr}\{(S - \delta\Sigma_o)Q_1SQ_2\} = (1-\delta)\,\mathrm{tr}(\Sigma_oQ_1\Sigma_oQ_2') + n^{-1}\{\mathrm{tr}(\Sigma_oQ_1\Sigma_oQ_2)$$

$$+ \mathrm{tr}(\Sigma_oQ_1)\,\mathrm{tr}(\Sigma_oQ_2')\} \quad . \tag{35}$$

Application of (35) to (34) then shows that

$$[\Omega]_{ij} = -\mathcal{E}\left(\frac{\partial^2 \log L_N}{\partial\gamma_i \partial\gamma_j}\bigg|_{\gamma=\gamma_o} \right)$$

$$= \frac{(n + p + 2)}{2} \operatorname{tr}\left(\frac{\partial\Sigma_o}{\partial\gamma_i} \Sigma_o^{-1} \frac{\partial\Sigma_o}{\partial\gamma_j} \Sigma_o^{-1} \right) + \frac{1}{2} \operatorname{tr}\left(\frac{\partial\Sigma_o}{\partial\gamma_i} \Sigma_o^{-1} \right) \operatorname{tr}\left(\frac{\partial\Sigma_o}{\partial\gamma_j} \Sigma_o^{-1} \right)$$

so that

$$\Omega = \frac{(n + p + 2)}{2} \Theta(\Sigma_o^{-1}) + \frac{1}{2} \Delta' \underset{-}{\sigma}_o \underset{-}{\sigma}_o' \Delta$$

and (33) follows. ||

In addition to yielding a B.G.L.S. estimator of γ_o, use of a consistent estimator of Σ_o^{-1} for V enables one to test the null hypothesis that (1) holds against the alternative that Σ_o is any positive definite matrix by means of the residual quadratic form $f(\hat{\gamma}|V)$.

Proposition 5. If $\bar{V} = \Sigma_o^{-1}$ and $\Sigma_o = \Sigma(\gamma_o)$, the limiting distribution of $nf(\hat{\gamma}|V) = 2^{-1}n \operatorname{tr}[\{S - \Sigma(\hat{\gamma})\}V]^2$ is chi-square with $p(p + 1)/2 - q$ degrees of freedom.

Proof. It was seen, using equation (32), that $\sqrt{n}(\hat{\gamma} - \ddot{\gamma})$ converges in probability to a null vector. Also $\sqrt{n}\{\underset{-}{\sigma}(\hat{\gamma}) - \underset{-}{\sigma}_o - \Delta(\hat{\gamma} - \gamma_o)\}$ converges in probability to a null vector since, by Taylor's theorem,

$$\sqrt{n}\,[\sigma_{ij}(\hat{\gamma}) - \{\sigma_{ij}(\gamma_o) + \frac{\partial\sigma_{ij}(\gamma_o)}{\partial\gamma'}(\hat{\gamma} - \gamma_o)\}]$$

$$= \frac{\sqrt{n}}{2}(\hat{\gamma} - \gamma_o)'\frac{\partial^2\sigma_{ij}(\gamma^*)}{\partial\gamma\partial\gamma'}(\hat{\gamma} - \gamma_o) \quad ,$$

where γ^* lies between $\hat{\gamma}$ and γ_o .

Consequently $\sqrt{n}\,\{s - \sigma(\hat{\gamma})\}$ converges stochastically to

$$\sqrt{n}\,[s - \sigma_o - \Delta(\ddot{\gamma} - \gamma_o)]$$

$$= \sqrt{n}\,[I - \Delta\{\Delta'(\Sigma_o^{-1} \otimes \Sigma_o^{-1})\Delta\}^{-1}\Delta'(\Sigma_o^{-1} \otimes \Sigma_o^{-1})](s - \sigma_o)$$

and

$$nf(\hat{\gamma}|V) = 2^{-1}n\{s - \sigma(\hat{\gamma})\}'(V \otimes V)\{s - \sigma(\hat{\gamma})\}$$

converges stochastically to

$$nf_o = 2^{-1}n(s - \sigma_o)'G_o(s - \sigma_o)$$

where

$$G_o = K_p^-\{(\Sigma_o^{-1} \otimes \Sigma_o^{-1}) - (\Sigma_o^{-1} \otimes \Sigma_o^{-1})\Delta[\Delta'(\Sigma_o^{-1} \otimes \Sigma_o^{-1})\Delta]^{-1}\Delta'(\Sigma_o^{-1} \otimes \Sigma_o^{-1})\}K_p^{-'} \quad .$$

Since $G_o\{K_p'(\Sigma_o \otimes \Sigma_o)K_p\}$ is idempotent of rank $\{p(p+1)/2 - q\}$ the limiting distribution of nf_o and of $nf(\hat{\gamma}|V)$ is the central chi-square distribution with $\{p(p+1)/2 - q\}$ degrees of freedom (Graybill, 1961, p. 83). ||

Anderson (1969, Section 4), considering linear covariance structures, has pointed out certain relationships between equations defining a G.L.S.

estimate with $V = \Sigma_o^{-1}$ and the Wishart likelihood equations. We shall now consider how, for covariance structures in general, an estimate of γ_o obtained by maximizing the Wishart likelihood function (M.W.L. estimator) may be regarded as a member of the class of B.G.L.S. estimates.

Proposition 6. Suppose that $\hat{\gamma}_1$ is a M.W.L. estimate of γ_o and that $\hat{\gamma}_2$ is a G.L.S. estimate where $V = \{\Sigma(\hat{\gamma}_1)\}^{-1}$. Then $\hat{\gamma}_2$ is a B.G.L.S. estimate and $\text{Prob}(\hat{\gamma}_1 \neq \hat{\gamma}_2) \to 0$ as $n \to \infty$.

Proof. Maximizing the Wishart likelihood function is equivalent to minimizing

$$F(\gamma) = \ln|\Sigma(\gamma)| - \ln|S| + \mathrm{tr}[S\{\Sigma(\gamma)\}^{-1}] - p \quad . \tag{36}$$

Consequently the equations,

$$\frac{\partial F(\gamma)}{\partial \gamma_i} = -\mathrm{tr}\{\Sigma^{-1}(S - \Sigma)\Sigma^{-1}\frac{\partial \Sigma}{\partial \gamma_i}\} = 0 \quad , \quad i = 1 \ldots q \quad , \tag{37}$$

and the condition that the matrix with typical element

$$\frac{\partial^2 F(\gamma)}{\partial \gamma_i \partial \gamma_j} = \mathrm{tr}\{\Sigma^{-1}(2S - \Sigma)\Sigma^{-1}\frac{\partial \Sigma}{\partial \gamma_i}\Sigma^{-1}\frac{\partial \Sigma}{\partial \gamma_j} - \Sigma^{-1}(S - \Sigma)\Sigma^{-1}\frac{\partial^2 \Sigma}{\partial \gamma_i \partial \gamma_j}\} \tag{38}$$

be positive definite will be satisfied at the point $\gamma = \hat{\gamma}_1$ ($\Sigma = \Sigma(\hat{\gamma}_1)$).

The equations

$$\frac{\partial f(\gamma|V)}{\partial \gamma_i} = -\mathrm{tr}\{V(S - \Sigma)V\frac{\partial \Sigma}{\partial \gamma_i}\} = 0 \quad , \quad i = 1 \ldots q \quad , \tag{39}$$

and the condition that the matrix with typical element

$$\frac{\partial^2 f(\underset{\sim}{\gamma}|V)}{\partial\gamma_i\partial\gamma_j} = \mathrm{tr}\{V \frac{\partial\Sigma}{\partial\gamma_i} V \frac{\partial\Sigma}{\partial\gamma_j} - V(S - \Sigma)V \frac{\partial^2\Sigma}{\partial\gamma_i\partial\gamma_j}\} \tag{40}$$

be positive definite will be satisfied at the point $\underset{\sim}{\gamma} = \hat{\underset{\sim}{\gamma}}_2$ when $V = \{\Sigma(\hat{\underset{\sim}{\gamma}}_1)\}^{-1}$.

Using similar reasoning to that used in the proof of Proposition 1 it can be shown (c.f. Anderson & Rubin, 1956, Theorem 12.1) that the M.W.L. estimator, $\hat{\underset{\sim}{\gamma}}_1$, is a consistent estimator of $\underset{\sim}{\gamma}_o$. Consequently $\{\Sigma(\hat{\underset{\sim}{\gamma}}_1)\}^{-1}$ is a consistent estimator of Σ_o^{-1} and $\hat{\underset{\sim}{\gamma}}_2$ is a B.G.L.S. estimator.

Equations (39) and (37) are equivalent when $V = \{\Sigma(\hat{\underset{\sim}{\gamma}}_1)\}^{-1}$. Consequently $\underset{\sim}{\gamma} = \hat{\underset{\sim}{\gamma}}_1$ is always a stationary point of $f(\underset{\sim}{\gamma}|\{\Sigma(\hat{\underset{\sim}{\gamma}}_1)\}^{-1})$ and will not be at a minimum only if the matrix with typical element given by (40) is not positive definite. Since the matrix with typical element (38) is positive definite at $\underset{\sim}{\gamma} = \hat{\underset{\sim}{\gamma}}_1$ and since the difference

$$\left[\frac{\partial^2 f(\underset{\sim}{\gamma}|\{\Sigma(\hat{\underset{\sim}{\gamma}}_1)\}^{-1})}{\partial\gamma_i\partial\gamma_j} - \frac{\partial^2 F(\underset{\sim}{\gamma})}{\partial\gamma_i\partial\gamma_j}\right]_{\underset{\sim}{\gamma}=\hat{\underset{\sim}{\gamma}}_1} = -2\ \mathrm{tr}\left[\Sigma^{-1}(S - \Sigma)\Sigma^{-1} \frac{\partial\Sigma}{\partial\gamma_i} \Sigma^{-1} \frac{\partial\Sigma}{\partial\gamma_j}\right]_{\underset{\sim}{\gamma}=\hat{\underset{\sim}{\gamma}}_1}$$

converges stochastically to zero, the probability that the matrix with typical element (40) is not positive definite at the point $\underset{\sim}{\gamma} = \hat{\underset{\sim}{\gamma}}_1$ tends to zero as $n \to \infty$. This implies that the probability that the point $\hat{\underset{\sim}{\gamma}}_1$ at which $F(\underset{\sim}{\gamma})$ has an absolute minimum does not give at least a relative minimum of $f(\underset{\sim}{\gamma}|\{\Sigma(\hat{\underset{\sim}{\gamma}}_1)\}^{-1})$ tends to zero as $n \to \infty$. Since $f(\underset{\sim}{\gamma}|\{\Sigma(\underset{\sim}{\gamma}_o)\}^{-1})$ is convex in a neighborhood of $\underset{\sim}{\gamma}_o$ and since $\hat{\underset{\sim}{\gamma}}_1$ and $\hat{\underset{\sim}{\gamma}}_2$ both converge stochastically to $\underset{\sim}{\gamma}_o$, the probability that there is a minimum at $\hat{\underset{\sim}{\gamma}}_1$ which does not coincide with the absolute minimum at $\hat{\underset{\sim}{\gamma}}_2$ tends to zero as $n \to \infty$. ||

This result implies that M.W.L. estimators will have the asymptotic properties of B.G.L.S. estimators provided only that the limiting distribution of S is the multivariate normal distribution specified earlier (and that the model satisfies the specified regularity conditions). No assumption of a Wishart distribution for S has been made.

Jöreskog & Goldberger (1972) have shown that the log likelihood ratio test statistic and a certain residual quadratic-form converge in probability in the particular case of unrestricted factor analysis. For covariance structures in general we may state:

Proposition 7. If $\hat{\gamma}$ is a B.G.L.S. (or M.W.L.) estimator, $nF(\hat{\gamma})$ and $nf(\hat{\gamma}|\{\Sigma(\hat{\gamma})\}^{-1})$ converge stochastically and have a limiting chi-square distribution with $p(p+1)/2 - q$ degrees of freedom.

Proof. Rearrangement of terms in (36) gives

$$F(\hat{\gamma}) = \mathrm{tr}\{(S - \hat{\Sigma})\hat{\Sigma}^{-1}\} - \ln|I + (S - \hat{\Sigma})\hat{\Sigma}^{-1}| \quad .$$

Using Taylor expansions in eigenvalues of $(S - \hat{\Sigma})\hat{\Sigma}^{-1}$, it is easily shown that

$$-\ln|I + (S - \hat{\Sigma})\hat{\Sigma}^{-1}| = \sum_{k=1}^{\infty} k^{-1}\,\mathrm{tr}\{-(S - \hat{\Sigma})\hat{\Sigma}^{-1}\}^k \quad .$$

Consequently,

$$nF(\hat{\gamma}) = nf(\hat{\gamma}|\hat{\Sigma}^{-1}) + n \sum_{k=3}^{\infty} k^{-1} \operatorname{tr}\{(\hat{\Sigma} - S)\hat{\Sigma}^{-1}\}^k$$

$$= nf(\hat{\gamma}|\hat{\Sigma}^{-1}) + o_p(1) \quad .$$

The limiting distribution of $nf(\hat{\gamma}|\hat{\Sigma}^{-1})$ follows from Proposition 5. ||

Consequently either $nf(\hat{\gamma}|\hat{\Sigma}^{-1})$ or $nF(\hat{\gamma})$ may be used in a large sample test of the null hypothesis that (1) holds when $\hat{\gamma}$ is a M.W.L. estimate. For many covariance structures the form of $F(\gamma)$ given in (36) simplifies at the minimum.

Proposition 8. Suppose that $\Sigma(\gamma)$ is such that, given any admissible $\hat{\gamma}$ and any positive scalar α, there is an admissible γ^* for which $\Sigma(\gamma^*) = \alpha\Sigma(\hat{\gamma})$. Then, if $\hat{\gamma}$ is a M.W.L. estimate, $\operatorname{tr}[S\hat{\Sigma}^{-1}] = p$ so that

$$F(\hat{\gamma}) = \ln|\hat{\Sigma}| - \ln|S| \quad .$$

This result was stated by Bock & Bargmann (1966, p. 521) for certain specific covariance structures. Their proof, however, applies to the general situation considered here.

4. Linear Covariance Structures

When $\Sigma(\gamma)$ is nonlinear, a successive approximation procedure, such as Newton's method, is required to obtain both G.L.S. and M.W.L. estimates. General expressions for the necessary derivatives are given in (37), (38), (39), and (40). When the specific forms of $\partial\Sigma/\partial\gamma_i$ and $\partial^2\Sigma/\partial\gamma_i\partial\gamma_j$ are

known, these expressions may be simplified using methods given by Bargmann (1967, Section 7).

When $\Sigma(\underset{\sim}{\gamma})$ is linear in $\underset{\sim}{\gamma}$, on the other hand, G.L.S. estimates may be expressed in closed form. A successive approximation procedure is still usually required for M.W.L. estimates (except in some special cases such as the compound symmetry model).

We can always express a linear structure $\Sigma(\underset{\sim}{\gamma})$ in the form

$$\underset{\sim}{\sigma}(\gamma) = \Delta\underset{\sim}{\gamma} \tag{41}$$

where Δ ($= \partial\underset{\sim}{\sigma}/\partial\underset{\sim}{\gamma}'$) is a known matrix of order $p^2 \times q$ and rank q.

Use of (39), (16), and (5) then shows that the G.L.S. estimates of $\underset{\sim}{\gamma}_o$ are:

$$\hat{\underset{\sim}{\gamma}} = \{\Theta(V)\}^{-1}\Delta' \operatorname{Vec}(VSV) \tag{42}$$

where

$$\Theta(V) = \Delta'(V \otimes V)\Delta \quad .$$

Whenever Δ is of full column rank and V is positive definite, $f(\underset{\sim}{\gamma}|V)$ is convex and has a unique minimum at $\underset{\sim}{\gamma} = \hat{\underset{\sim}{\gamma}}$. $\Theta(V)$ then is positive definite.

If V is a fixed matrix (e.g., $V = I$), or a stochastic matrix distributed independently of S, $\hat{\underset{\sim}{\gamma}}$ is an unbiased estimator of $\underset{\sim}{\gamma}_o$. If V is a consistent estimator of Σ_o^{-1} (e.g., $V = S^{-1}$ or $V = \{\Sigma(\hat{\underset{\sim}{\gamma}})\}^{-1}$), $\hat{\underset{\sim}{\gamma}}$ is a B.G.L.S. estimator of $\underset{\sim}{\gamma}_o$ and $2n^{-1}\{\Theta(V)\}^{-1}$ is a consistent

estimator of the asymptotic covariance matrix of $\hat{\gamma}$ (Proposition 3). Also, the statistic

$$nf(\hat{\gamma}|V) = 2^{-1}n\ \mathrm{tr}[(S - \hat{\Sigma})V(S - \hat{\Sigma})V]$$

$$= 2^{-1}n[s'(V \otimes V)s - \hat{\gamma}'\{\Theta(V)\}\hat{\gamma}]$$

is approximately distributed as chi-square with $\{p(p + 1)/2\} - q$ degrees of freedom if n is large and the null hypothesis $\sigma_0 = \Delta\gamma_0$ holds (Proposition 5).

The M.W.L. $\hat{\gamma}$ is defined by (42) with V replaced by $\{\Sigma(\hat{\gamma})\}^{-1}$, and will simultaneously be a G.L.S. estimate in the sense of minimizing $f(\gamma|\{\Sigma(\hat{\gamma})\}^{-1})$ (Proposition 6) whenever $\Sigma(\hat{\gamma})$ is positive definite. This M.W.L. estimate may be calculated by means of a successive approximation procedure:

1./ Use (42) with $V = S^{-1}$ to obtain $\hat{\gamma}_{(1)}$.

2./ Use (42) with $V = \{\Sigma(\hat{\gamma}_{(1)})\}^{-1}$ to obtain $\hat{\gamma}_{(2)}$.

3./ Continue in this way until the differences $\hat{\gamma}_{(i+1)} - \hat{\gamma}_{(i)}$ become sufficiently small.

It is easily shown that this successive G.L.S. procedure is equivalent to the Fisher scoring method (Kendall & Stuart, 1967, pp. 48-49) for obtaining M.W.L. estimates. (When $\Sigma(\gamma)$ is not linear in γ, however, minimizing $f(\gamma|\{\Sigma(\hat{\gamma}_{(i)})\}^{-1})$ to obtain $\hat{\gamma}_{(i+1)}$ is no longer equivalent to the Fisher scoring method.)

The successive G.L.S. estimators $\hat{\gamma}_{(1)}, \hat{\gamma}_{(2)}, \hat{\gamma}_{(3)} \cdots$ are all B.G.L.S. estimators and have the same asymptotic properties. It is therefore difficult to justify the calculation of precise M.W.L. estimates, particularly if more than three or four iterations are required.

McDonald (1972) has investigated patterned covariance structures where subsets of elements of Σ are equal or have a known value, usually zero. In such models, where elements of Δ are either 1 or 0, (42) would be employed without further algebraic manipulation to provide G.L.S. estimates. Use of (4) would avoid storage of the large matrix $V \otimes V$ by a computer program.

In other linear covariance structures, however, Δ involves direct products of certain matrices and (42) may be simplified considerably. We shall now examine such models in greater detail. They are of the form

$$\Sigma = A\Phi A' + D_{\psi} \tag{43}$$

where the $p \times m$ "model matrix" A is known and of full column rank, Φ is symmetric of order M, and D_{ψ} is diagonal of order p. Models of this kind have been discussed by Bock & Bargmann (1966, p. 510), Mukherjee (1970), and Jöreskog (1970a, Sections 2.4 and 2.5). Newton methods for obtaining M.W.L. estimates of Φ_o and D_{ψ_o} are available (Bock & Bargmann, 1966; Anderson, 1970) and the methods proposed by Jöreskog (1970a) may also be employed.

It will be convenient to consider separately the cases where Φ is diagonal, $\Phi = D_{\phi}$, and where Φ is symmetric but not diagonal.

Case I. Φ is diagonal.

When $\Phi = D_\phi$, (43) may be expressed in the form of (41) with

$$\Delta = \{(A \otimes A)H_m, H_p\} \quad ,$$

$$\underline{\gamma}' = (\underline{\phi}', \underline{\psi}') = \{\text{diag}'(D_\phi), \text{diag}'(D_\psi)\} \quad ,$$

$$q = m + p \quad .$$

Then, using (15), it can be shown that

$$\Theta(V) = \begin{pmatrix} (A'VA)*(A'VA) & (A'V)*(A'V) \\ (VA)*(VA) & V*V \end{pmatrix} \quad . \tag{44}$$

and, using (5) and (14), that

$$\Delta' \text{ Vec}(VSV) = \begin{pmatrix} \text{diag}(A'VSVA) \\ \text{diag}(VSV) \end{pmatrix} \quad . \tag{45}$$

Substitution of (44) and (45) in (42) now provides the estimate $\hat{\underline{\gamma}}$. The matrix to be inverted, $\Theta(V)$, is positive semidefinite provided that V is positive definite. Singularity of the matrix implies that $\underline{\gamma}_o$ is not identified.

We have minimized $f(\underline{\gamma}|V)$ without imposing any constraints and some elements of $\hat{\underline{\gamma}}$ could be negative. The elements of $\underline{\gamma}_o' = (\underline{\phi}_o', \underline{\psi}_o')$, however, represent variances (cf. Bock & Bargmann, 1966) so that it would be preferable for the elements of $\hat{\underline{\gamma}}$ to be nonnegative. Minimization of $f(\underline{\gamma}|V)$ subject to the inequality constraints

$$\hat{\gamma}_i \geq 0 \quad , \quad i = 1 \dots q \tag{46}$$

ERIC

may be accomplished by applying the "sweep" operator (Dempster, 1969, Section 4.3.2; Morgan & Tatar, 1972) to the symmetric matrix Q of order $q + 1$ which is defined initially as

$$Q = \begin{pmatrix} Q_{11} & \underline{q}_{12} \\ \underline{q}_{12}' & q_{22} \end{pmatrix} .$$

where

$$Q_{11} = \Theta(V) \text{ as defined in (44)} ,$$

$$\underline{q}_{12} = \Delta' \operatorname{Vec}(VSV) \text{ as defined in (45)} ,$$

$$q_{22} = \underline{s}'(V \otimes V)\underline{s} = \operatorname{tr}(VSVS) . \tag{47}$$

The superscript * will be used to indicate that the sweep operator has been applied on a particular row of Q. An element of $\underline{q}_{12}$, $[\underline{q}_{12}]_i^*$, lies in row i^* of Q on which the sweep operator has been applied. Applying the reverse sweep operator on the same row of Q cancels the sweep operation so that $[\underline{q}_{12}]_i^*$ becomes $[\underline{q}_{12}]_i$.

The minimization algorithm is:

1/ Sweep Q on row i if $[\underline{q}_{12}]_i \geq 0$:

$$[\underline{q}_{12}]_i \rightarrow [\underline{q}_{12}]_i^* \geq 0 .$$

2/ If 1/ results in a $[\underline{q}_{12}]_j^*$, in a row $j^* \neq i^*$ on which Q has previously been swept, becoming negative, reverse sweep Q on row j^* :

$$[\underline{q}_{12}]_j^* \rightarrow [\underline{q}_{12}]_j < 0 .$$

5./ Continue until all $[\underline{q}_{12}]^*_i \geq 0$ and all $[\underline{q}_{12}]_i < 0$, i or i* $\leq$ m + p . The sweep operator is never applied on the last row of Q .

Then $\hat{\underline{\gamma}}$ is given by

$$\hat{\gamma}_i = [\underline{q}_{12}]^*_i \quad , \text{ if } Q \text{ has been swept on row } i = i^*$$

$$= 0 \quad , \text{ if } Q \text{ has not been swept on row } i$$

and $nf(\hat{\underline{\gamma}}|V)$ may be obtained from

$$nf(\hat{\underline{\gamma}}|V) = \frac{n}{2} q_{22} \quad .$$

Since

$$\left.\frac{\partial f(\underline{\gamma}|V)}{\partial \gamma_i}\right|_{\underline{\gamma}=\hat{\underline{\gamma}}} = -n[\underline{q}_{12}]_i \quad , \text{ if } Q \text{ has not been swept on row } i$$

$$= 0 \quad , \text{ if } Q \text{ has been swept on row } i = i^* \quad ,$$

the Kuhn-Tucker conditions are satisfied,

$$\hat{\gamma}_i \geq 0$$

$$\frac{\partial f(\hat{\underline{\gamma}}|V)}{\partial \gamma_i} \geq 0$$

$$\hat{\gamma}_i \cdot \frac{\partial f(\hat{\underline{\gamma}}|V)}{\partial \gamma_i} = 0$$

and $\hat{\underline{\gamma}}$ is a global minimum of $f(\underline{\gamma}|V)$ subject to the inequality constraints (46) (Fiacco & McCormick, 1968, pp. 89-90).

The sweep operator may then be applied on the remaining rows of Q_{11} (where $[\underline{q}_{12}]_i < 0$) to obtain $\{\Theta(V)\}^{-1}$.

In some cases some elements of $\underset{\sim}{\gamma}_o$ may be in known ratio. For example, suppose that

$$D_{\psi_o} = \psi_o D_\alpha$$

where D_α is a known diagonal matrix (e.g., $D_\alpha = I$). Then $\underset{\sim}{\gamma}_o' = (\underset{=}{\phi}_o', \psi_o)$, $q = m + 1$, and estimates are obtained as before with

$$q_{11} = \begin{pmatrix} (A'VA)*(A'VA) & \{(A'V)*(A'V)\}\underset{=}{\alpha} \\ \underset{=}{\alpha}'\{(VA)*(VA)\} & \underset{=}{\alpha}'(V*V)\underset{=}{\alpha} \end{pmatrix}$$

$$q_{12} = \begin{pmatrix} \mathrm{diag}(A'VSVA) \\ \underset{=}{\alpha}'\mathrm{diag}(VSV) \end{pmatrix}$$

and q_{22} defined by (47) .

Similar procedures may be employed when other elements of γ_o are equal or in known ratio.

Case II. Φ is symmetric.

In (41) we now have

$$\Delta = \{(A \otimes A)K_m^{-'} , H_p\} \quad .$$

$$\gamma' = (\underset{\sim}{\phi}', \underset{=}{\psi}') \quad .$$

$$q = \{ \quad . + 1)/2\} + p \quad .$$

After some algebra, making use of the methods of Section 2, (42) can be simplified to:

$$\hat{\Phi} = B'(S - \hat{D}_\psi)B \tag{48}$$

$$\hat{\underline{\underline{\psi}}} = W \operatorname{diag}[VSV - GSG] \tag{49}$$

where

$$B = VA(A'VA)^{-1}$$

$$G = VA(A'VA)^{-1}A'V$$

$$W = (V*V - G*G)^{-1} \quad .$$

The matrix to be inverted to give W is a submatrix of $(V + G) \otimes (V - G)$ and is therefore positive semidefinite provided that V is positive definite. Singularity of the matrix implies that $\underline{\gamma}_o$ is not identified.

It is of interest to note that, although the number of parameters to be estimated in Case II is greater than that in Case I, the largest matrix to be inverted in (48), (49) is of order p while the inversion of a matrix of order $(p + m)$ is required when (44), (45), (42) are employed.

Taking $Q_{11} = (V*V - G*G)$, $\underline{q}_{12} = \operatorname{diag}[VSV - GSG]$, and $q_{22} = \operatorname{tr}[VSVS - GSGS]$ and replacing $\hat{\underline{\gamma}}$ by $\hat{\underline{\underline{\psi}}}$, the algorithm described under Case I may be employed to give a $\hat{\underline{\underline{\psi}}}$ satisfying the inequality constraints

$$\hat{\psi}_{ii} \geq 0 \quad , \qquad i = 1 \ldots p \quad . \tag{50}$$

When $\hat{\underline{\underline{\psi}}}$ has been obtained, $\hat{\Phi}$ may be obtained from (48). This gives the absolute minimum of $f(\underline{\gamma}|V)$ subject to the inequality constraints (50). It is possible that $\hat{\Phi}$, an estimated dispersion matrix, will not be positive semidefinite. To ensure that $\hat{\Phi}$ is positive semidefinite one

could replace Φ by TT' , but the model would then no longer be linear and the estimates would be more difficult to obtain.

If V is a consistent estimator of Σ_o^{-1} , we have

$$\widehat{\mathrm{Cov}}(\hat{\gamma},\hat{\gamma}') = 2n^{-1}\{\Theta(V)\}^{-1} ,$$

with elements:

$$\widehat{\mathrm{Cov}}(\hat{\psi}_i,\hat{\psi}_j) = 2n^{-1}w_{ij} ,$$

$$\widehat{\mathrm{Cov}}(\hat{\phi}_{ij},\hat{\psi}_k) = -2n^{-1}\sum_{r=1}^{p} b_{ri}b_{rj}w_{rk} ,$$

$$\widehat{\mathrm{Cov}}(\hat{\phi}_{ij},\hat{\phi}_{gh}) = n^{-1}(c_{ig}c_{jh} + c_{ih}c_{jg} + 2\sum_{r=1}^{p}\sum_{s=1}^{p} b_{ri}b_{rj}w_{rs}b_{sg}b_{sh}) ,$$

where

$$c_{ig} = [(A'VA)^{-1}]_{ig} .$$

The case where the elements of $\underline{\underline{\psi}}_o$ are in known ratio,

$$D_{\psi_o} = \psi_o D_\alpha$$

may be treated as in Case I. Taking

$$w = \{\underline{\underline{\alpha}}'(V{*}V - G{*}G)\underline{\underline{\alpha}}\}^{-1}$$

we have:

$$q = \{m(m + 1)/2\} + 1 ,$$

$$\hat{\psi} = w\underline{\underline{\alpha}}' \operatorname{diag}(VSV - GSG) ,$$

$$\hat{\Phi} = B'(S - \hat{\Psi}D_\alpha)B \quad ,$$

$$\widehat{\mathrm{Var}}(\hat{\Psi}) \;.\; 2n^{-1}w \quad ,$$

$$\widehat{\mathrm{Cov}}(\hat{\phi}_{ij}, \hat{\Psi}) = -2n^{-1}w[B'D_\alpha B]_{ij} \quad ,$$

$$\widehat{\mathrm{Cov}}(\hat{\phi}_{ij}, \hat{\phi}_{gh}) = n^{-1}(c_{ig}c_{jh} + c_{ih}c_{jg} + [B'D_\alpha B]_{ij}[B'D_\alpha B]_{gh}) \quad .$$

Formulae, both in Case I and Case II, simplify in an obvious manner when $V = S^{-1}$. When maximum likelihood estimates are being obtained and $V = (A\hat{\Phi}A' + \hat{D}_\psi)^{-1}$ the following well-known identities may be employed to reduce computation if $|\hat{D}_\psi| \neq 0$:

$$V = \hat{D}_\psi^{-1} - \hat{D}_\psi^{-1}A(\hat{\Phi}^{-1} + A'\hat{D}_\psi^{-1}A)^{-1}A'\hat{D}_\psi^{-1} \quad ,$$

$$(A'VA)^{-1}A'V = (A'\hat{D}_\psi^{-1}A)^{-1}A'\hat{D}_\psi^{-1} \quad .$$

We note, also, that Proposition 8 applies in both Case I and Case II.

The Fisher scoring algorithm employed here for obtaining M.W.L. estimates may require more iterations to attain convergence than existing Newton algorithms, but less computation is required during each iteration. This reduction in computation per iteration is particularly noticeable in Case II.

The B.G.L.S. estimates obtained using S^{-1} for V require less computation than the M.W.L. estimates and have the same desirable asymptotic properties. Small sample properties of the estimators are as yet unknown. In a Monte Carlo experiment (Durand, 1971) use of S^{-1} for V gave

estimates which appeared more biased ($\mathcal{E}(\hat{\underset{\sim}{\gamma}}) < \underset{\sim}{\gamma}_0$) than the M.W.L. estimates but which, however, appeared to be as precise in terms of mean squared error of estimation. Also, in practical applications of both Case I and Case II procedures, the author has observed that taking $V \quad S^{-1}$ tends to give estimates which are slightly smaller than the M.W.L. estimates. A similar tendency in factor analysis was noted by Jöreskog & Goldberger (1972) .

This tendency is apparent in the example given in Table 1a. It shows G.L.S. estimates ($V = I$, $V = S^{-1}$) and M.W.L. ($V = \hat{\Sigma}^{-1}$) estimates of parameters in a quasi-simplex model based on a covariance matrix obtained by Bilodeau (1957) in a study of a two-hand coordination task. This matrix has been reported by and analyzed by Bock & Bargmann (1966) and by Jöreskog (1970b). The model is:

$$\Sigma = AD_{\phi}A' + \psi I$$

where

$$a_{ij} = 1 \quad , \qquad p \geq i \geq j \geq 1$$

$$= 0 \quad , \qquad i < j \quad .$$

It can be seen that the G.L.S. estimates with $V = S^{-1}$ and the M.W.L. estimates ($V = \hat{\Sigma}^{-1}$) agree rather closely and differ somewhat from the unweighted least squares estimates ($V = I$).

The successive G.L.S. (Fisher scoring) algorithm for obtaining M.W.L. estimates converged to four figures on the third iteration. Estimates of standard errors and values of the test statistics are given in Tables 1b and 1c.

Acknowledgments

I am grateful to Ingram Olkin, Karl G. Jöreskog and Leon J. Gleser for helpful comments on an earlier version of this paper.

Table 1. Bilodeau's Example.

a) Estimates of parameters in a quasi simplex model.

V	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\phi}_4$	$\hat{\phi}_5$	$\hat{\phi}_6$	$\hat{\psi}$
I	504.1	63.3	31.1	124.6	36.7	22.7	19.3
S^{-1}	452.3	53.4	15.4	74.4	20.6	0.0	44.3
$\hat{\Sigma}^{-1}$	482.6	54.6	15.9	81.4	21.6	1.5	45.3

b) Estimates of standard errors. $\text{diag}^{\frac{1}{2}}\{76\Theta(V)\}^{-1}$.

V	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\phi}_4$	$\hat{\phi}_5$	$\hat{\phi}_6$	$\hat{\psi}$
S^{-1}	56.9	14.6	10.2	14.5	9.5	10.1	4.8
$\hat{\Sigma}^{-1}$	58.7	14.6	10.2	14.9	9.6	10.2	4.7

c) Test statistics. d.f. = 14 . n = 152 .

V	$nf(\hat{\underline{\gamma}}\|V)$	$nF(\hat{\underline{\gamma}})$
S^{-1}	9.34	
$\hat{\Sigma}^{-1}$	9.24	9.46

References

Anderson, T. W. (1969). Statistical inference for covariance matrices with linear structure. In Multivariate Analysis, vol. II, ed. P. R. Krishnaiah, pp. 55-66. New York: Academic Press.

Anderson, T. W. (1970). Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices. In Essays in Probability and Statistics, eds. R. C. Bose, et al., pp. 1-24. Chapel Hill: University of North Carolina Press.

Anderson, T. W. & Rubin, H. (1956). Statistical inference in factor analysis. In Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. V, ed. J. Neyman, pp. 111-150. Berkeley: University of California Press.

Bargmann, R. E. (1967). Matrices and determinants. In Chemical Rubber Company Handbook of Tables for Mathematics, ed. S. M. Selby, pp. 144-66. Cleveland: Chemical Rubber Company.

Beckenbach, E. F. & Bellman, R. (1965). Inequalities. New York: Springer-Verlag.

Bilodeau, E. A. (1957). The relationship between a relatively complex motor skill and its components. Am. J. Psychol. 70, 49-55.

Bock, R. D. & Bargmann, R. E. (1966). Analysis of covariance structures. Psychometrika 31, 507-34.

Cramér, H. (1946). Mathematical Methods of Statistics. Princeton, N.J.: Princeton University Press.

Dempster, A. P. (1969). Elements of Continuous Multivariate Analysis. Reading, Mass.: Addison-Wesley.

Durand, F. D. (1971). Monte Carlo estimates for covariance structures. Unpublished B.Sc. (Hons.) Thesis. Durban: University of Natal.

Fiacco, A. V. & McCormick, G. P. (1968). Nonlinear Programming: Sequential Unconstrained Minimization Techniques. New York: Wiley.

Graybill, F. A. (1961). An Introduction to Linear Statistical Models New York: McGraw-Hill.

Jöreskog, K. G. (1970a). A general method for analysis of covariance structures. Biometrika 57, 239-51.

Jöreskog, K. G. (1970b). Estimation and testing of simplex models. Brit. J. Math. Statist. Psychol. 23, 121-45.

Jöreskog, K. G. & Goldberger, A. S. (1972). Factor analysis by generalized least squares. Psychometrika 37, 243-60.

Kendall, M. G. & Stuart, A. (1967). The Advanced Theory of Statistics, vol. 2, second edition. London: Griffin.

Kendall, M. G. & Stuart, A. (1969). The Advanced Theory of Statistics, vol. 1, third edition. London: Griffin.

McDonald, R. P. (1972). Testing pattern hypotheses for covariance matrices. Psychometrika (In press).

Morgan, J. A. & Tatar, J. F. (1972). Calculation of the residual sum of squares for all possible regressions. Technometrics 14, 317-25.

Mukherjee, B. N. (1970). Likelihood ratio tests of statistical hypotheses associated with patterned covariance matrices. Brit. J. Math. Statist. Psychol. 23, 89-120.

ERIC

Searle, S. R. (1966). Matrix Algebra for the Biological Sciences. New York: Wiley.

ERIC